CALCULATING CHANGE IN REGIONAL ACCESSIBILITY DUE TO AUTONOMOUS VEHICLES

Spandana Anand

Capstone Project (CP 6950)
Georgia Institute of Technology

Abstract –

The following project tries to answer the question "How will autonomous vehicles affect growth in the Metro Atlanta region?" We attempt to do this through calculating how accessibility will change based on traffic conditions. We also determine how it compares to the predicted population/employment growth by the Atlanta Regional Commission and the kind of land use patterns that are present in those regions.

Introduction –

The future of vehicle technology looks very promising with the rise of automated technology. But the implications it carries are hard to understand and even more difficult to quantify. One of the major advantages of having autonomous vehicles would be reduced travel time due to an increased traffic flow. It would also mean increased access to places that took too long to reach before.

The aim of this project is to model a traffic vs. no traffic situation to see what regions we can expect growth to happen in the future due to increased access. We'll also compare that against demographic and land use patterns.

Review of academic literature –

Not only is the technology of autonomous vehicles advancing rapidly, it is expected to become commonplace in the next 15 years (III, A. H., 2018). Depending on how they are regulated, we could see denser cities as extra parking lots turn into residential or commercial districts (The Economist, 2018).

Even a few automated vehicles on the road can reduce fuel consumption and excessive braking, helping alleviate traffic (Leong, J., 2018). In that case, we would definitely have faster access to places that previously had a time barrier, and more space for development as land patterns start to change.

Methodology –

The analysis consisted of various steps that required the use of multiple softwares in order to achieve the final results. The whole method has been described in this section step by step.

Gathering vector data -

The vector datasets required that were readily available are as described below:

- Census tract boundaries and relevant data – The analysis has been conducted at the census tract level since it seemed like the best scale considering the final analysis parameter, i.e., land use. The polygon shapefile was obtained from ARC's website and has demographic data forecasted until the year 2040
- Roads TIGER file – This was downloaded from the official US Census Bureau's website for the 21- county region. This would be our main dataset since network analysis is the primary technique used in the project.
- Land Pro 2012 – The polygon shapefile for this was also downloaded from ARC's website. It contains a column with the different land use classes as specified by them
- County Boundaries – This was just to clip other files to the required extent

Clipping data -

The extent of the project includes the main 9 counties of the metro Atlanta region, namely Fulton, DeKalb, Gwinnett, Cobb, Clayton, Coweta, Douglas, Fayette, and Henry counties. This seemed appropriate as the size of census tracts in the selected area are more homogeneous than in the counties located further out.

The datasets clipped and their methods are described below –

- Census tracts – Using 'Select by location' where the centroids of the feature fell inside the boundaries
- Roads – Using 'Select by location' where the feature intersected the source feature layer. This was done to ensure no connecting roads that could be used to traverse the region were left out

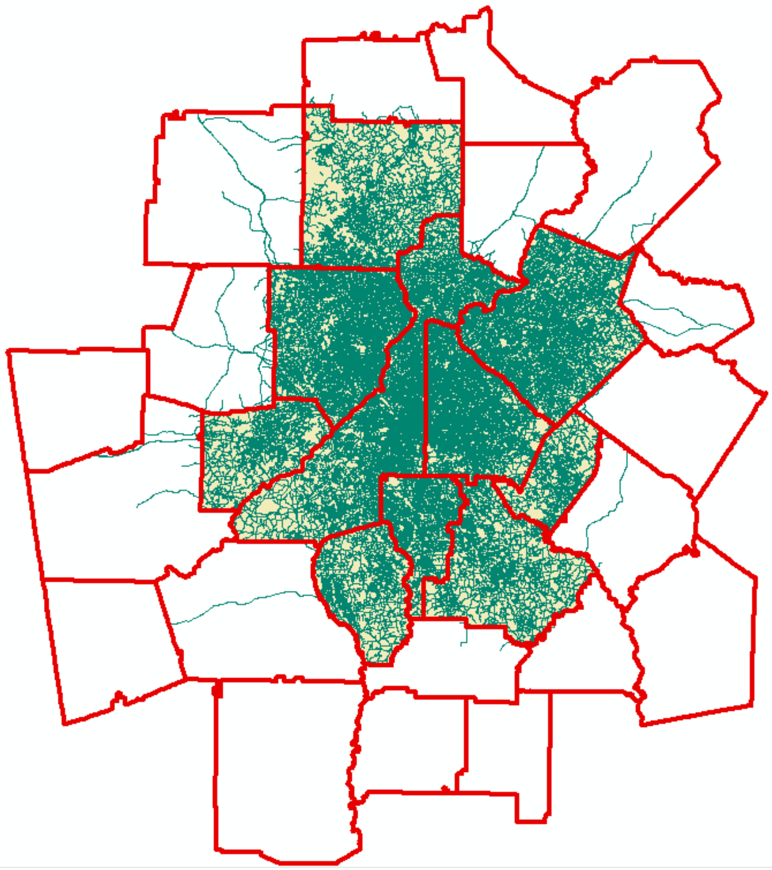


Figure 1 – Clipping the data

The land use data did not need to be clipped as we will be using intersect on it later

Obtaining traffic data -

The main aim of the project is to calculate change in accessibility due to decrease in traffic caused by autonomous vehicles becoming more mainstream. To do this, it was necessary to recognize the areas that are heavily affected by traffic right now.

The source used for downloading traffic information was Mapquest (https://www.mapquest.com/) using the API that they give when you sign up for their service. They have live traffic information available which is available to download as an image. An example photo is as shown below.

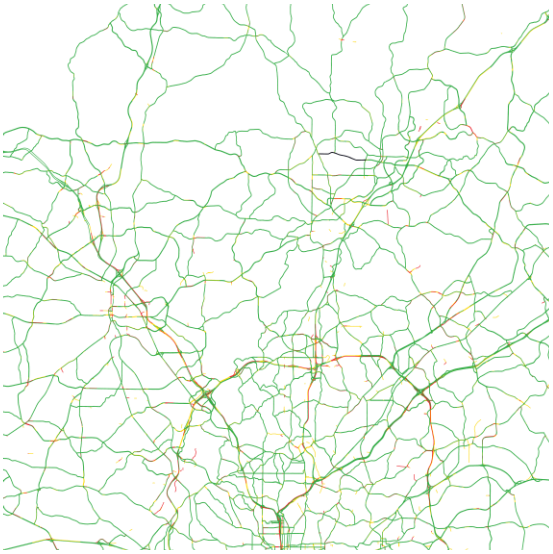


Figure 2 - Example of traffic flow grid

The traffic data was obtained during peak hours (around 5:15 pm). The whole extent of our data was divided into grids using the ‘Grid index features’ tool, then the coordinates of their centroids were used to obtain the photos of traffic flow to maximize the number of roads we’d obtain the data for. The total no. of grids came out to be 14 and the photos were stitched together using photoshop to come up with the final image as shown below.

Figure 3 - Traffic flow for the project region

Calculating travel times –

Before building a network, it was necessary to figure out where and by how much the travel times would change depending on the traffic. Two different travel times were calculated as explained below.

Future scenario - No traffic assumed:

This was done using Field Calculator after calculating the length of each line segment in miles. The average speed is assumed to be 60 miles/hour and the formula used is

$$\text{Time} = \text{Distance}/\text{Speed}$$

This was then multiplied by 60 to obtain the time taken to traverse each line segment in minutes.

Present Scenario – with peak traffic:

The traffic data was used to calculate a factor with which to multiply the times calculated in the previous section. But since it was in a .gif format, it had to be edited to fit our data. This was done in several stages.

Georeferencing the image - The georeferencing toolbar available in ArcMap was used to fit the image to the roads dataset. A total of 5 control points were used since the alignment did not need to be very accurate.

Figure 4 - Georeferencing the traffic flow image to roads

Converting to tiff –

Since the gif format does not allow for proper conversion between raster and polygon, the image was converted into tiff to better access the traffic data that we had just obtained.

Recognizing areas of medium and high traffic (yellow and red) – The traffic flow data is usually represented with three colors, i.e, red, yellow and green. The converted raster (tiff) file had three bands (referred to as band 1, 2 and 3 respectively) which had their own values for each of the different areas on the image. A table created using sample pixels from different areas on the image gave the following values:

Area of image	*Band 1*	*Band 2*	*Band 3*
Green	30	163	50
Yellow	254	214	0
Red	232	14	32
White space	255	255	255

Table 1 - Average values of sample pixels

Separating the bands -

Since band 3 showed very minimal difference in values between the three areas of traffic, it was discarded. The other two bands were separated into two layers using the 'Make Raster Layer' tool and then classified. An image of the two bands with corresponding traffic flow is shown below.

Figure 5 - Traffic flow values for the two bands mapped

Reclassifying the rasters –

The first band was reclassified into 8 different categories using an increment of 5 due to the difference in red and yellow of traffic was less than 20. The second band was classified using natural breaks (Jenks) and 9 classes since the range of values for different areas of the image was much further apart. The reclassified rasters were given the following values:

BAND 1			BAND 2		
S.NO.	Pixel value	Classification	**S.no.**	Pixel value	Classification
1	0 - 220	Green	**1**	0 - 25	Red
2	220 - 225	Green	**2**	25 - 48	Red
3	225 - 230	Green	**3**	48 - 74	Red
4	230 - 235	Red	**4**	74 - 105	Green
5	235 - 240	Red	**5**	105 - 140	Green
6	240 - 245	Yellow	**6**	140 - 176	Green
7	245 - 250	Yellow	**7**	176 - 211	Green
8	250 - 254	Yellow	**8**	211 - 241	Yellow
9	255	NoData	**9**	241 - 244	NoData
			10	255	NoData

Table 2 - Reclassifying raster

Converting to polygon and combining -

In order to intersect the traffic data with the roads, they both needed to be in the same format. This was accomplished by:

- Converting raster to polygon: The reclassified rasters were converted using 'Raster to polygon' tool. The resulting polygon layers contained every individual pixel with a grid value (values provided during classification)
- Dissolving: The polygon files were dissolved using the grid ID's as the value to dissolve on
- Subset data: The polygons recognized by the values described above were separated into layers for the two levels of traffic condition (red and yellow)
- Intersect: The resulting layers for the two levels obtained from the two bands gave different areas in their output with some overlap. The intersecting areas was assumed to be more accurate and this was extracted, giving us our final layers to be combined with the roads dataset.
- The image below shows a sample area of the two traffic layers.

Figure 6 - Final polygons representing traffic flow

Intersecting with roads -

The final step before creating a network was to calculate a new travel time variable for the roads which would take into account the traffic condition. This was done by running field calculator on roads selected using 'Select by location' using a distance buffer of 20 meters from the traffic layers. Any roads that were selected in both the cases were assumed to be high traffic (red). The factors used to multiply the normal time variable was 1.5 for medium traffic and 2 for high traffic.

Creating a network and calculating access –

Two different networks were built using the network analyst extension to calculate the accessibility value of census tracts: one with traffic and one without.

Creating origins and destinations:

Since network analyst uses point locations to for its calculations, a layer was created for the centroid of each census tract. This was done by first calculating the x and y coordinates of the tracts, then making an XY event layer using these coordinates, and finally exporting it as a layer.

The centroids did not line up exactly with the road edges so the 'Snap (Editing)' tool was used to move the points to intersect with the closest available road in the network. An example of this operation has been shown in the image below.

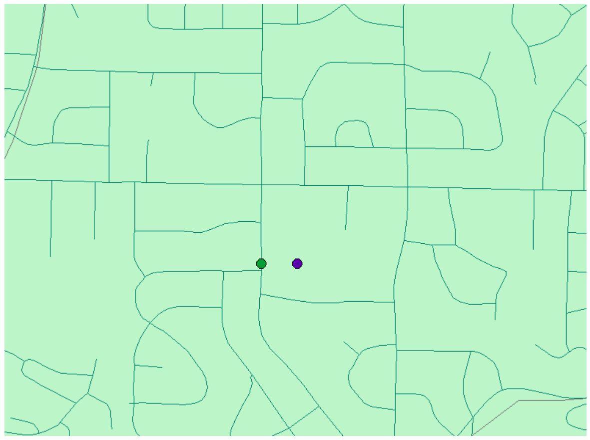


Figure 7 - Example of a snapped census tract centroid

The OD cost matrix:

Using the centroids of the census tracts as both origin and destination points, an OD cost matrix was generated with a time limit of 30 mins, which is around the average commute time in Atlanta (Pendered, D., 2017). The network with traffic generated 180,490 lines, while the one without generated 234,244 lines, indicating that there are going to be more connections between census tract if there was no traffic, which is the condition assumed with autonomous vehicles. The corresponding line layers were exported for further analysis.

Calculating access:

Since the OD cost matrix exports a concatenated string of origin and destination IDs, it had to be split to calculated access per census tract. This was done using field calculator and using the python code [Name].split(" ")[0], which returned the first word of the string, i.e., the IDs of the points of origin.

The layers were then dissolved using the newly calculated Origin ID as the base and ObjectID count as statistic to get the total number of census tracts that are accessible by each of them. Both of these values were joined back to the Census tract shapefile and a new field was calculated to show increase in accessibility.

Calculating land use for each of the census tracts –

After calculating which areas will see increased access due to the decrease in traffic, the next step was to figure out the land use patterns of the census tracts to compare.

Intersecting data:

The Land Pro data was combined with the census tracts using the 'intersect' tool. This would give us the individual polygons of every land use category corresponding to each of the census tracts. A new field was added to recalculate the number of land use classes to six, namely:

- Residential low density
- Residential medium density
- Residential high density
- Residential multi
- Commercial
- Other

Another field was added to calculate the area of each of these polygons in acres and the layer was dissolved using the Tract ID and new land classification, with a statistic being used to sum the area.

Converting to wide format:

The new layer did contain areas for the new land classification corresponding to each census tract, but these were nestled together under one variable. They had to be converted so each of these land use classes could be represented individually for each of the census tracts.

This was achieved by exporting the attribute table of the layer as an excel file and manipulating it in R using the spread command from the dplyr library. The resulting table was exported as csv and then manipulated further in R using a pivot table to compress the Census tract IDs to individual records and sum the areas for each corresponding land use class. An example of a few records before and after manipulation has been shown below.

OBJECTID	STFID	LandUse	SUM_Area
1	13057090100	COMMERCIAL	342.6766582
2	13057090100	OTHER	37228.6349
3	13057090100	RES_HIGH	72.28439441
4	13057090100	RES_LOW	5852.042465
5	13057090100	RES_MED	844.8041578
6	13057090200	COMMERCIAL	97.61418579
7	13057090200	OTHER	47677.8757
8	13057090200	RES_HIGH	39.85159906
9	13057090200	RES_LOW	6518.589727
10	13057090200	RES_MED	569.2070871
11	13057090200	RES_MULTI	6.940553299

Table 3 - Land use data in long form

Row Labels	Sum of COMMERCIAL	Sum of OTHER	Sum of RES_HIGH	Sum of RES_LOW	Sum of RES_MED	Sum of RES_MULTI
13057090100	342.6766582	37228.6349	72.28439441	5852.042465	844.8041578	0
13057090200	97.61418579	47677.8757	39.85159906	6518.589727	569.2070871	6.940553299
13057090300	28.51543902	24458.4004	260.1505721	3329.872832	1727.83439	21.67898637
13057090400	594.7708919	8413.669444	153.5784938	1494.220178	633.9627763	141.0719401

Table 4 - Land use data converted to wide form

Joining back to census tract:

The final table from the above step was imported into ArcMap and joined back to the census tract layer file to calculate the percentage of area covered by each land use type by adding new fields for each and using the total acres of each census tract.

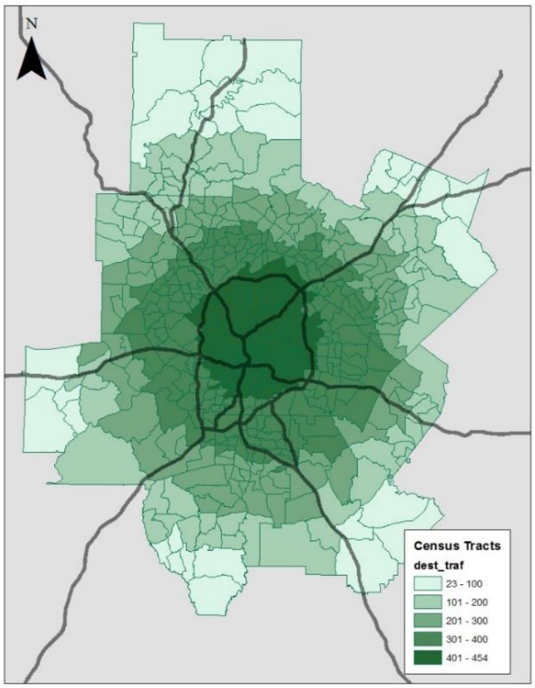
CTs_ARC.Area	CT_ID	CTs_ARC.Perc_comm	CTs_ARC.Perc_oth	CTs_ARC.Perc_low	CTs_ARC.Perc_med	CTs_ARC.Perc_high
44340.447879	13057090100	0.772831	83.960891	13.197978	1.905268	0.163021
54888.569489	13057090200	0.177841	86.863032	11.876042	1.037023	0.072605
29826.453329	13057090300	0.095605	82.002376	11.16416	5.79296	0.872214
11431.274606	13057090400	5.203015	73.602199	13.071335	5.545863	1.343494
23847.558541	13057090501	0.421207	80.959673	15.17329	3.445841	0
20084.942398	13057090502	0.813788	59.842302	31.639673	7.041221	0.663056
5420.54421	13057090601	5.672709	44.225962	39.148655	7.509126	2.479714

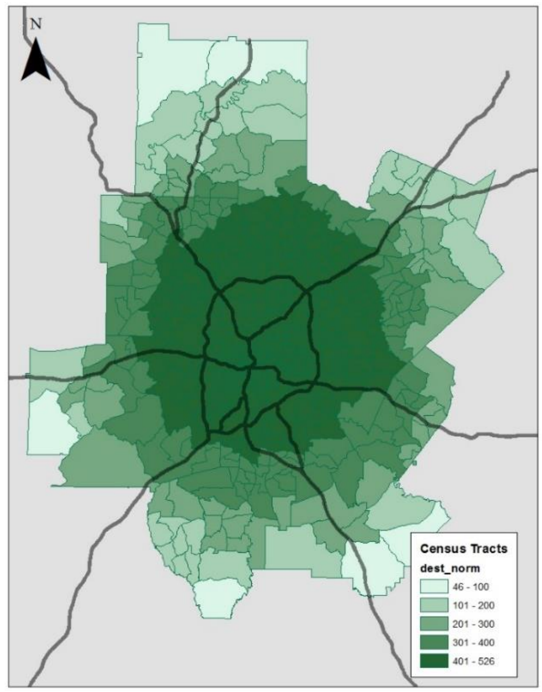
Figure 8 - Final attribute table for the census tract feature class

Results –

Change in accessibility:

The images below show the accessibility with traffic and without, and the change in access for the census tracts in terms of both sheer number and percentage.





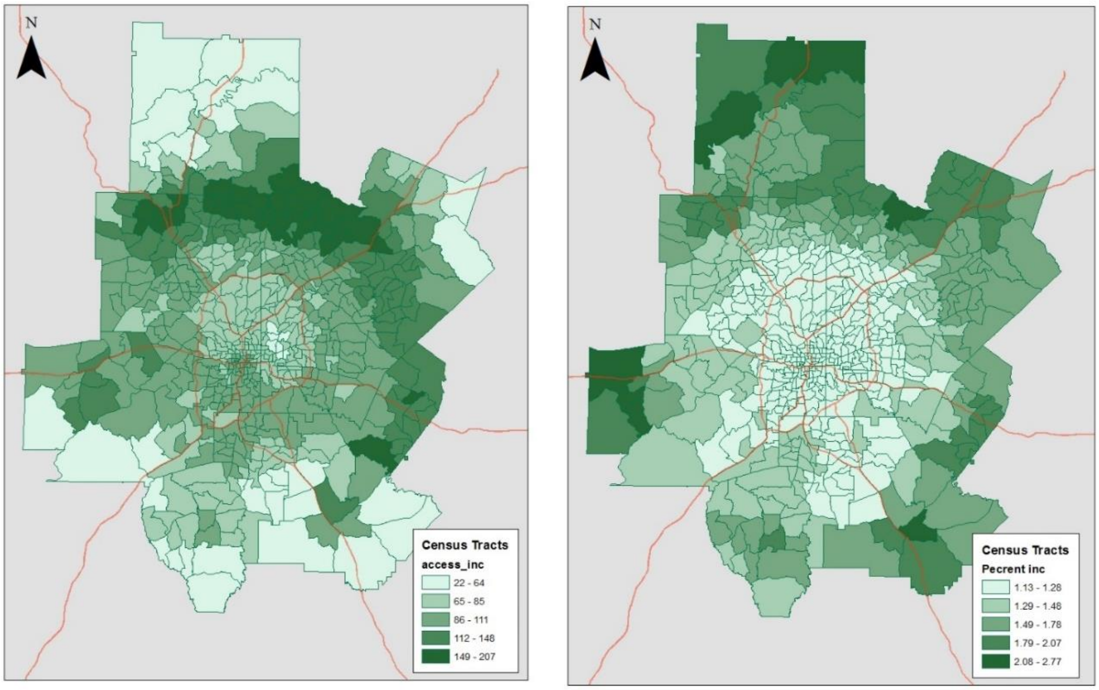


Figure 9 - Accessibility before and after for the metro Atlanta region

The two images on top show access of each census tract with and without traffic. As we can see, access increases in almost all the regions shown, and it seems to be proportional. But in order to get into more detail, we can use the bottom image.

The bottom left image shows increase in accessibility by number of census tracts within a half hour drive. We can see that most of this increase is focused on the northern regions of Fulton, Gwinnett and Cobb counties, with a few regions showing increase in the southwest region. This leads us to believe that the population density will now be capable of moving further north. The second image, which shows a percentage increase in accessibility (without traffic/with traffic) shows growth almost at the edges of our project boundary, but again, focused more on the northern region.

Another image below shows the increase in access per acre.

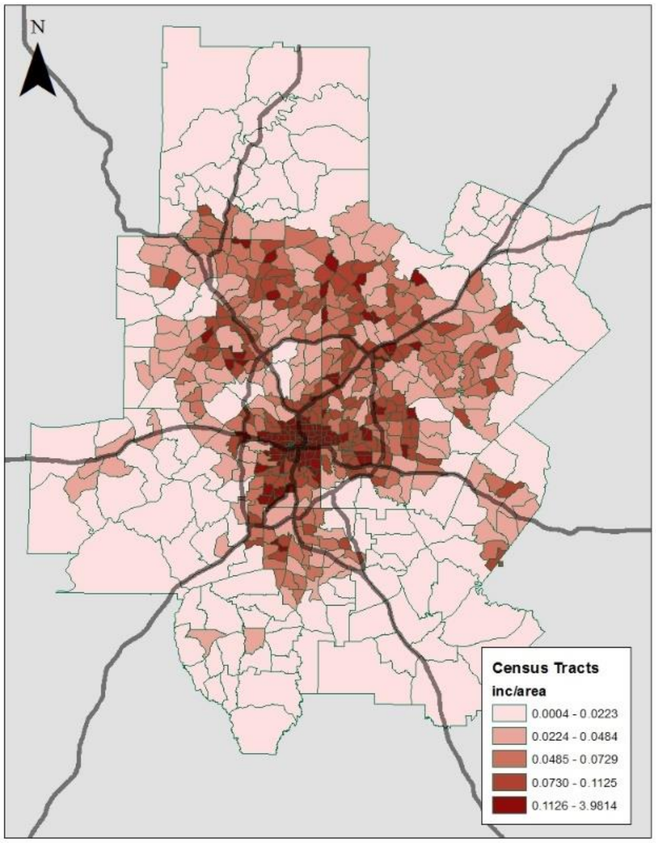


Figure 10 - Increase in accessibility per acre

This gives us a much clearer idea of what's going on. When combined with the previous analysis, the northern Fulton and Cobb county regions definitely seem to be the areas that can expect the most amount of growth due to increased access.

Change in population and employment

ARC's data has predictions for change in demographic data starting from 2016 going up to 2040. We'll look at some of these changes through the images shown below.

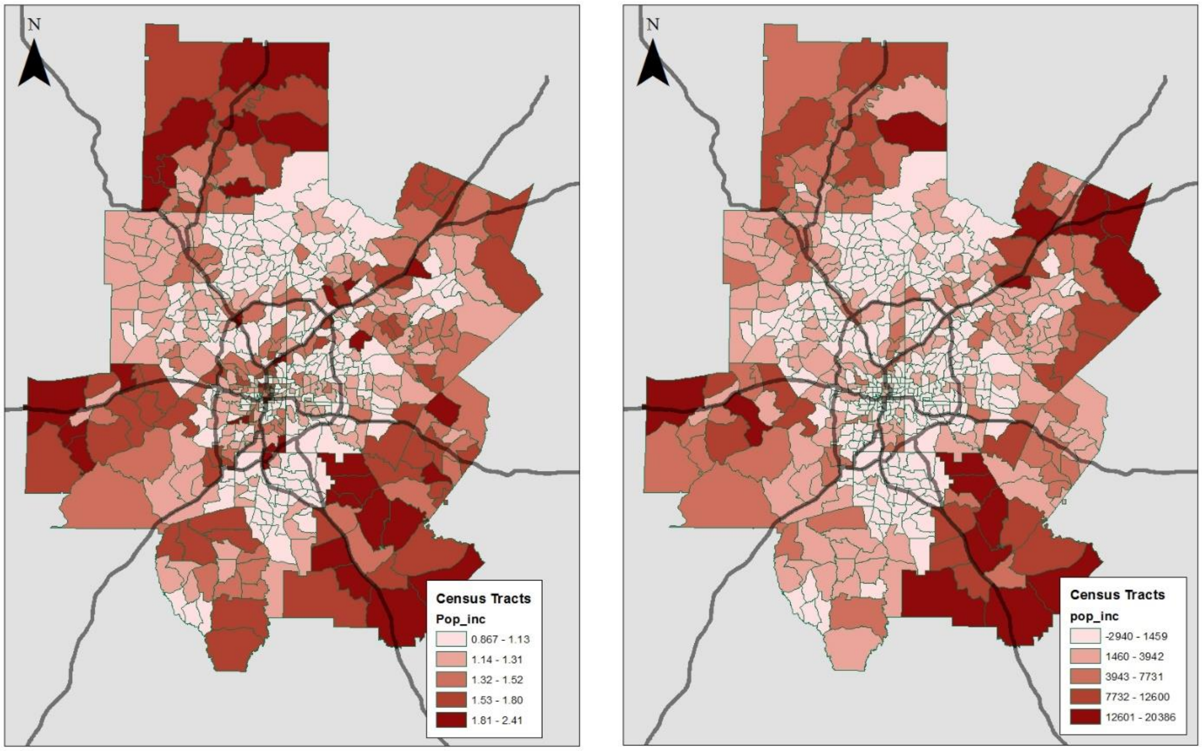


Figure 11 - Population change from 2016 to 2040

The left image represents percentage population growth and the right total change in population from the year 2016 to 2040 (predicted). We see that the growth pattern here does correspond to the percentage increase in access, but not the absolute value or growth per acre, which would be much better indicators. This shows a discrepancy between what is expected by ARC and what may happen due to the increased number of autonomous vehicles on the road.

The images below show the same calculations as above but for employment instead.

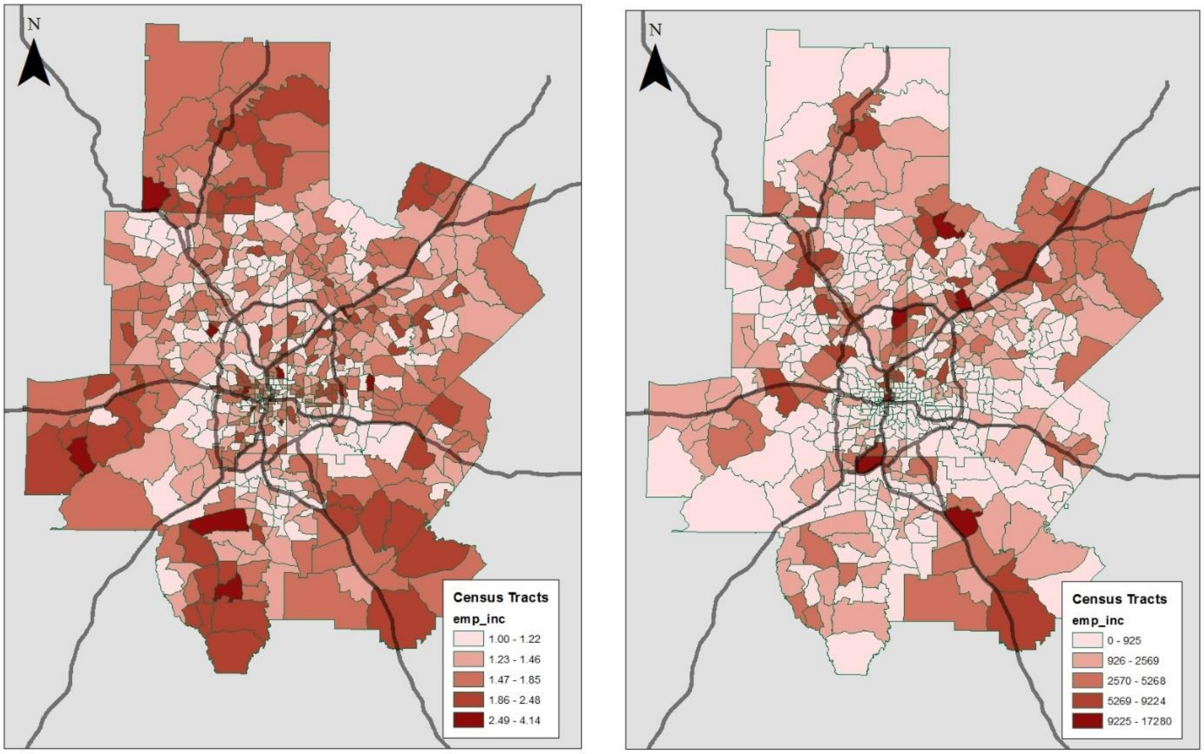


Figure 12 - Employment change from 2016 to 2040

The employment growth is a lot more evenly distributed throughout the region, both in terms of total and percentage change. We do see some amount of grouping in the north and north eastern regions. This area is close to where the population growth is expected due to increased access.

Combining all three variables, i.e., population, employment an access, gives us a better idea of where we can expect to see increased sprawl in the future.

Current land use

Now that we have some areas delimited for future growth, we're going to study the land use composition of the region. The two main categories studied are residential and commercial land use.

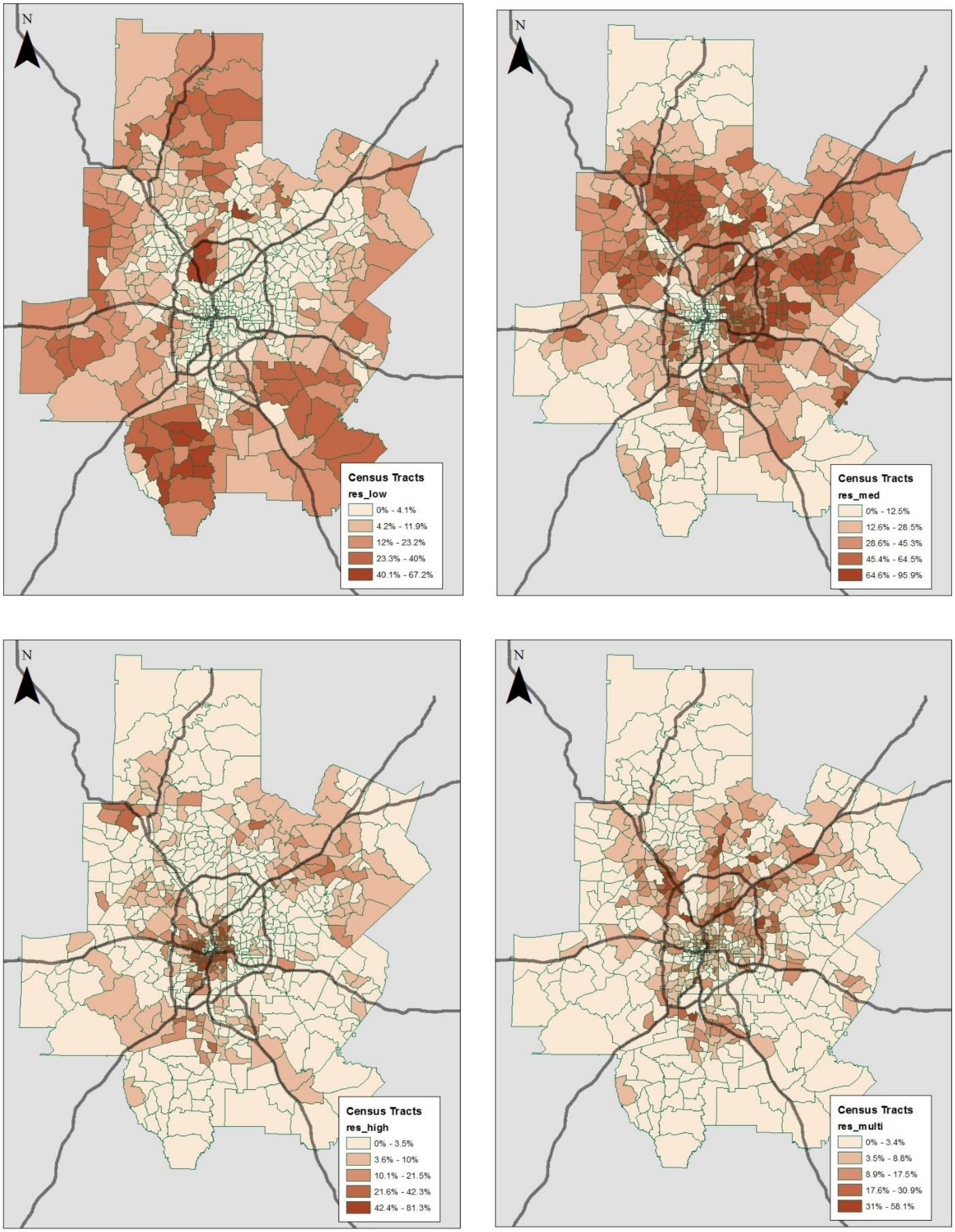


Figure 13 - Percentage area covered by each category of residential density

The images above show the percentage of each kind of residential property in every census tract, namely low density, medium density, high density and multi family. Of the areas that were observed previously to expect growth, the northern Fulton and Gwinnett counties mostly have medium density residential properties, with some high and low density mixed in. Cobb county does show more of a tendency to have low density residences.

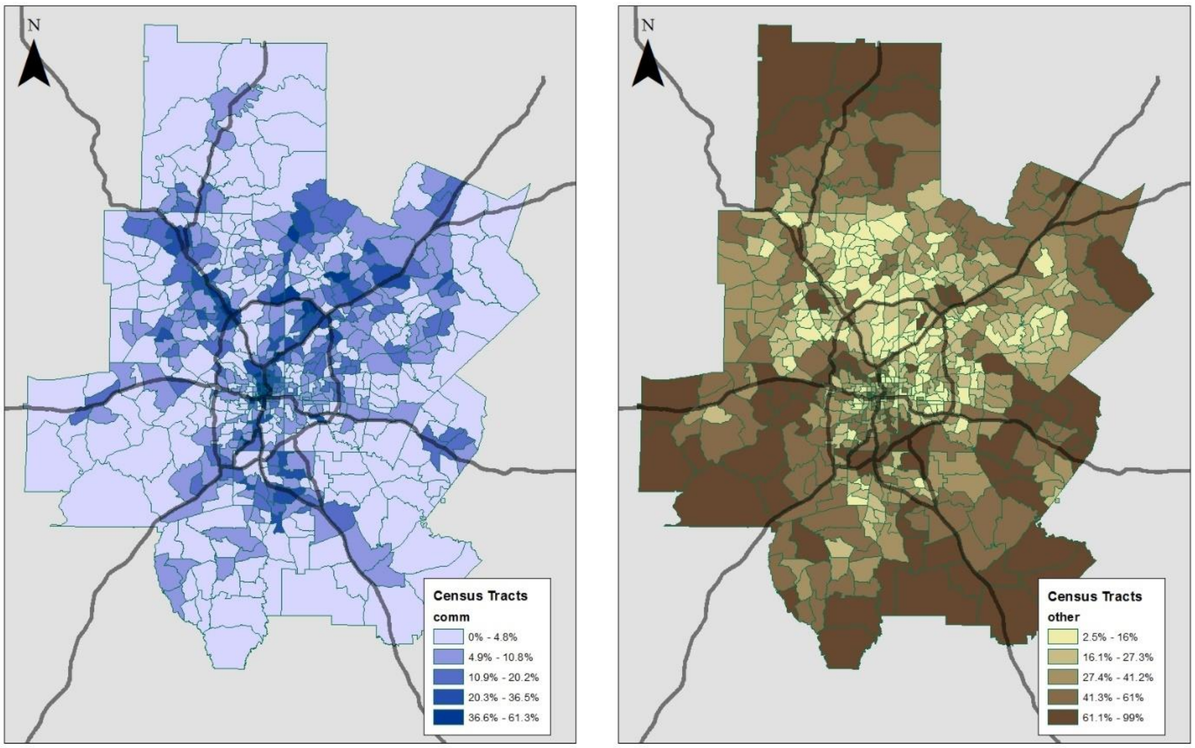


Figure 14 - Percentage area covered by commercial and other land uses

The above two images show percentage of commercial and other areas in the project region to help provide further information. A lot of the areas discussed above seem to have a heavy amount of commercial activity as well. This provides the policy makers and governmental agencies a good opportunity to control and direct the growth that may happen in those areas in the near future.

Significance –

With the rapid advancement in technology, we will soon begin to see autonomous vehicles on the roads. But the kind of changes that this will cause on the land use pattern of the city is still unknown. This project was an attempt at recognizing which areas we could possibly see the growth in and, therefore, be able to regulate or direct it better when it does happen.

The overall structure of the model works, but there are a few aspects that could still be refined. Using the traffic factors in a gravity model to calculate travel times may have arguably produced more accurate results. As would building an accessibility index instead of just creating a count of how many census tracts are served by each of them in the two scenarios.

In conclusion, we do see the results that the model was supposed to show. There are even some anomalies discovered when we compare the expected increase in access to where ARC predicts the growth will happen. We can apply the same logic to other metro regions, and we can also add as many variables as necessary to better predict what we can expect to see in the future and enact policies accordingly.

Additional notes:

An attempt was made to calculate normal travel time using Google API after obtaining coordinates of end points of each line segment in the road dataset using ArcPy, but this was unsuccessful due to the number of records being too high (103083) since Google has a daily limit of 2500 records.

An accessibility index could not be calculated since the export tool exceeded the maximum amount of records allowed (around 65000). Hence, a normal dissolve operation was used.

References –

III, A. H. (2018, June 04). Most people expect driverless cars to become common, and they worry about it. Retrieved June 22, 2018, from https://www.washingtonpost.com/local/trafficandcommuting/most-people-expect-driverless-cars-to-become-common-and-they-worry-about-it/2018/06/03/fa213ea0-64ed-11e8-a768-ed043e33f1dc_story.html?noredirect=on&utm_term=.2404edf695ee

Leong, J. (2018, February 20). Study shows autonomous vehicles can help improve traffic flow. Retrieved June 22, 2018, from https://phys.org/news/2018-02-autonomous-vehicles-traffic.html

Pendered, D. (2017, December 10). Census says commute time in metro Atlanta up 30 seconds in seven years, transit usage dips. Retrieved June 22, 2018, from https://saportareport.com/census-says-commute-time-metro-atlanta-30-seconds-seven-years-transit-usage-dips/

Who Is Behind the Wheel? Autonomous Vehicles. (2018, March 3). *The Economist (US)*.